Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Quantifying the overall effect of biotic interactions on species distributions along environmental gradients^{*}

Marc Ohlmann ^{a,b,*}, Catherine Matias ^c, Giovanni Poggiato ^{a,d}, Stéphane Dray ^{b,e}, Wilfried Thuiller ^a, Vincent Miele ^{b,e}

^a Univ. Grenoble Alpes, CNRS, Univ. Savoie Mont Blanc, LECA, Laboratoire d'Ecologie Alpine, F-38000 Grenoble, France

^b CNRS, UMR5558, Laboratoire de Biométrie et Biologie Évolutive, F-69622 Villeurbanne, France

^c Sorbonne Université, Université Paris Cité, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, F-75005 Paris, France

^d Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, F-38000 Grenoble, France

^e Université de Lyon, F-69000 Lyon, France

ARTICLE INFO

Keywords: Biodiversity patterns C-score Environmental niche Markov random fields Metanetwork Species co-occurrence

$A \ B \ S \ T \ R \ A \ C \ T$

Separating environmental effects from those of interspecific interactions on species distributions has always been a central objective of community ecology. Despite years of effort in analysing patterns of species cooccurrences and the developments of sophisticated tools, we are still unable to address this major objective. A key reason is that the wealth of ecological knowledge is not sufficiently harnessed in current statistical models, notably the knowledge on interspecific interactions.

Here, we develop ELGRIN, a statistical model that simultaneously combines knowledge on interspecific interactions (*i.e.*, the metanetwork), environmental data and species occurrences to tease apart their relative effects on species distributions. Instead of focusing on single effects of pairwise species interactions, which have little sense in complex communities, ELGRIN contrasts the overall effect of species interactions to that of the environment.

Using various simulated and empirical data, we demonstrate the suitability of ELGRIN to address the objectives for various types of interspecific interactions like mutualism, competition and trophic interactions. We then apply the model on vertebrate trophic networks in the European Alps to map the effect of biotic interactions on species distributions. We find that altitude, species richness and connectance significantly impact the overall effect of biotic interactions.

Data on ecological networks are everyday increasing and we believe the time is ripe to mobilise these data to better understand biodiversity patterns. ELGRIN provides this opportunity to unravel how interspecific interactions actually influence species distributions.

1. Introduction

Ecologists have always strived to understand the drivers of biodiversity patterns with the particular interest to tease apart the effects of environment and biotic interactions on species distributions and communities (Ricklefs, 2008; Thuiller et al., 2015; Chase and Leibold,

2003; de Candolle, 1855). Species distributions are influenced by the abiotic environment (e.g. climate or soil properties) because of their own physiological constraints that allow them or not to sustain viable populations in specific environmental configurations (Austin, 2002; Pulliam, 2000). However, the occurrence of a species in a given site

* Corresponding author.

E-mail address: marcohlmann@live.fr (M. Ohlmann).

https://doi.org/10.1016/j.ecolmodel.2023.110424

Received 9 January 2023; Received in revised form 24 May 2023; Accepted 28 May 2023 Available online 14 June 2023 0304-3800/© 2023 Elsevier B.V. All rights reserved.







^A **Open Research statement:** ELGRIN is implemented in the function elgrin of the R package econetwork available on the code repository https: //plmlab.math.cnrs.fr/econetproject/econetwork and at CRAN (https://cran.r-project.org/). The simulation procedure can be reproduced with the Rmarkdown vignettes available along with this manuscript and on the code repository. The vertebrate data from O'Connor et al. (2020) can be found at https:// datadryad.org/stash/dataset/doi:10.5061/dryad.bcc2fqz79. Climatic data were downloaded from the Worldclim v2 database (http://www.worldclim.org/bioclim) as described in the Methods, section 2.4. Land cover data were downloaded from Global Land cover v2.2 (http://due.esrin.esa.int/page_globcover.php), net primary productivity was downloaded from (https://sedac.ciesin.columbia.edu/data/set/hanpp-net-primary-productivity/data-download) and the human footprint index was downloaded from http://sedac.ciesin.columbia.edu/data/set/wildareas-v2-human-footprint-geographic, searching for the latest version (V2 the time of the article). Pre-processed data together with analysis scripts are available on the code repository.

is also influenced by other species through all sort of interactions that can be trophic (e.g. a predator needs preys), non-trophic (e.g. plant species need to be pollinated by insects) or competitive (two species with the same requirements might exclude each other) (Guisan et al., 2017; Gravel et al., 2019; Lortie et al., 2004; Soberón and Nakamura, 2009).

Teasing apart the effects of environmental variations and interspecific interactions on species distributions and communities from observed co-occurrence patterns has always been a hot topic in ecology since the earlier debate between Diamond (1975) and Connor and Simberloff (1979), to the recent syntheses on the subject (Blanchet et al., 2020). More than anything, with a few exceptions, and despite recent advances like joint species distribution models (Ovaskainen et al., 2017) or null model developments (Peres-Neto et al., 2001; Chalmandrier et al., 2013), the conclusion has been that it is almost impossible to retrieve and estimate interspecific interactions from observed spatial patterns of species communities (Zurell et al., 2018; Blanchet et al., 2020). This conclusion should thus preclude any attempt to disentangle the relative effects of environment and interspecific interactions. A major difficulty of this long-standing issue is that interspecific interactions could be of any type (i.e. positive, negative, asymmetric) and that observed patterns average out all these interactions. Observed communities indeed reflect the overall outcome of interspecific interactions that is difficult to dissect, especially when analysing pairwise species spatial associations as it is commonly done (e.g., Tikhonov et al., 2017) . Yet, this overall outcome might be worth analysing on its own, for instance to measure the overall strength of interspecific interactions in a given community and between communities, how it depends on the co-existing species, and how it varies in space.

Interestingly, so far there have been few attempts to integrate the wealth of existing knowledge to address this fundamental ecological issue (Blanchet et al., 2020; Holt, 2020). Indeed, the spatial analysis of biotic interactions is gaining an increased interest with novel technologies to measure interactions in the field (e.g. camera-traps, gutcontent), open databases (e.g. GLOBI, Mangal) and the developments of new statistical tools to analyse them (Tylianakis and Morris, 2017; Pellissier et al., 2018; Ohlmann et al., 2019; Botella et al., 2022). The combination of expert knowledge, literature, available databases, and phylogenetic hypotheses has also given rise to large metanetworks that generalise the regional species-pool of community ecology by incorporating the potential interactions between species from different trophic levels along with their functional and phylogenetic characteristics (Maiorano et al., 2020; Morales-Castilla et al., 2015). Despite a few attempts (e.g., Staniczenko et al., 2017), information on interaction networks has been poorly integrated to understand and model biodiversity patterns. We believe that the time is ripe to incorporate network information into the process of modelling species distributions and communities. It implies to integrate both biotic and abiotic information (and their spatial variations) as explanatory factors in statistical models to weight their relative strength.

In this article, we propose a novel statistical model, called ELGRIN (in reference to Charles Elton and Joseph Grinnell) that can handle the effects of both environmental factors and known interspecific interactions (aka a metanetwork) on species distributions. We rely on Markov random fields (MRF, also called Gibbs distribution, e.g., Brémaud, 1999), a family of flexible models that can handle dependencies between variables using a graph. More specifically, ELGRIN jointly models the presence and absence of all species in a given area in function of environmental covariates and the topological structure of the known metanetwork (Fig. 1 left). It separates the interspecific interaction effects (Fig. 1 top-right) from those of the environment (Fig. 1 bottom-right) on species distributions. To our knowledge, EL-GRIN is the first model whose outputs are the relative strengths of biotic factors needed on top of abiotic environmental variables to shape the species distributions and their spatial variation (see Latitude/Longitude in Fig. 1 top-right). It thus provides a convenient way to integrate network ecology in joint species community modelling.

In this article, we first present the overall modelling framework and then assess its performances under different scenarios implying data simulated using three different dynamic models. In other words, although ELGRIN considers only static observational data (metaweb and community data), we evaluated the model using simulated data generated using different dynamic models that involve various underlying processes, including intraspecific competition. We test the ability of ELGRIN to decipher the relative importance of abiotic and interspecific interactions in these difficult cases so as to better understand what kind of signal ELGRIN can or cannot retrieve from the data. Finally, we apply the model on vertebrate trophic networks in the European Alps as an empirical study.

2. Material and methods

2.1. Species data and potential interactions

We consider a set of sites or locations indexed by $l \in \{1, ..., L\}$, where we observe the occurrence (presence/absence) of *N* species and also a set of environmental variables (vector W_i).

For the same set of N species, we assume that we know all the pairwise interactions between them (e.g. who eats whom), an information summarised with a graph $G^{\star} = (V^{\star}, E^{\star})$ over the set of nodes $V^{\star} =$ $\{1, \ldots, N\}$ and edges E^{\star} . This graph, usually called a metanetwork, summarises a regional pool of both species and interactions and can be obtained by aggregating local networks at different locations or by using expert knowledge and literature review (e.g., Cirtwill et al., 2019; Maiorano et al., 2020). Note that various types of interactions can be considered here (e.g., trophic, mutualism, competition). However, while considering various interaction types is technically possible, the interpretation of results would be difficult because in our framework, G^{\star} records the presence of an interaction and not its type. An additional note is that our model, like most species community models (e.g. Joint species distribution models, ordination techniques) relying on occurrence data, makes some assumptions about the ecological processes structuring species assemblages. In our current implementation of ELGRIN, we consider that only unimodal responses of species to environmental gradients and interspecific interactions shape communities, ignoring other processes such as dispersal limitation or mass effect for instance. Lastly, note also that our model supposes that the graph associated to the metanetwork is undirected with no self-loops (see model specifications below) and thus ignores intraspecific interactions. Hereafter, we refer to co-present (or co-absent) species, pairs of species that are connected in the metanetwork and jointly present (or absent, respectively) at a given location.

2.2. The statistical model of ELGRIN

Model description The aim of ELGRIN model is to factorise the joint species presence distribution between a Grinnellian part, that consists in a regression on environmental covariables, and an Eltonian part that quantifies association strengths between species distribution according to the metanetwork. More formally, we consider a set of random variables $\{X_i^l\}_{i \in V^*}$ taking values in $\{0, 1\}$ and that represent the presence/absence of species $i \in V^*$ at location $l \in \{1, \dots, L\}$. We rely on a *Markov random field* (see for instance Brémaud, 1999) to model the dependencies between species occurrences at location l. This is a multivariate model that encodes statistical dependencies between species distribution using a network. In our ELGRIN model, these dependencies are encoded through the metanetwork G^* . For each location $l \in \{1, \dots, L\}$, we thus assume that these random variables are distributed according to a Gibbs distribution specifying the joint



Fig. 1. Schematic view of ELGRIN statistical model. Given (a) an interaction metanetwork that summarises known interactions (edges) between species (nodes), (b) species occurrences data and (c) environmental covariates for a set of sites, ELGRIN model estimates (d) the overall effect of known biotic interactions on species distributions in each site using two association parameters, and (e) the environmental response of each species along all sites using regression parameters on environmental covariates.

associations between the species occurrence variables $\{X_i^l\}_{i\in V^\star}$, as follows:

$$\mathbb{P}(\{X_i^l\}_{i \in V^{\star}}) = \frac{1}{Z} \exp\left(\sum_{i \in V^{\star}} [a_l + a_i + W_l^{\mathsf{T}} b_i + (W_l^2)^{\mathsf{T}} c_i] X_i^l \right)$$
(1a)

+
$$\beta_{l,co-pres} \sum_{(i,j)\in E^{\star}} 1\{X_j^l = X_i^l = 1\}$$
 (1b)

+
$$\beta_{l,co-abs} \sum_{(i,j)\in E^{\star}} \mathbf{1}\{X_{j}^{l} = X_{i}^{l} = 0\}$$
 (1c)

where $1{A}$ is the indicator function of event *A* (either co-absence $X_j^l = X_i^l = 0$ or co-presence $X_j^l = X_i^l = 1$), notation U^{T} stands for the transpose of vector *U* and *Z* a normalising constant discussed below. Some model parameters have an ecological interpretation (Table 1). The use of W_l and W_l^2 (the vector of coordinate-wise squared values of W_l) allows modelling a quadratic species response to environmental gradient, following then a bell-shaped relationship as expected under classical niche theory (Chase and Leibold, 2003).

Sub-Eq. (1a) is the Grinnellian part of ELGRIN, as it represents some prior probability of species occurrences independently of their interactions. Parameters a_i, b_i, c_i capture the response of species *i* to environment, seen through a vector of environmental covariates W_i . The intercepts a_i and a_i are estimated up to a constant only (see Appendix S1: Section S.2.1) and may not be interpreted, whereas the vectors b_i, c_i deal with the species environmental niche, like in a standard species distribution model (Guisan et al., 2017).

Sub-Eqs. (1b) and (1c) form the Eltonian part of ELGRIN. It considers only interactions $(i, j) \in E^*$, i.e. the edges of the metanetwork. The β_l represent the overall influence of the interactions (as encoded through G^*) on all species presence/absence at location *l*. However, this influence may be different for co-presence and co-absence, with parameters $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ respectively (see Table 2). When a $\beta_{l,co-pres}$ is positive, it represents a positive driving force of co-presence on species distributions. By contrast, a negative value indicates that species co-presences are avoided. The same reasoning holds

with $\beta_{l,co-abs}$ for co-absences. Since the interaction parameter $\beta_{l,co-abs}$ can also be influenced by co-absences between species that are both absent at location *l* only because of unsuitable environmental conditions, we introduced a compatibility matrix so that the effect of interactions is only estimated in the environmental conditions where interacting species could co-occur (details are given in Appendix S1: Section S.2.2). Importantly, this compatibility matrix is estimated during the inference procedure and is not a required input by the user.

Note that we chose the parameters β_l to be specific to location $l \in \{1, ..., L\}$ such that the effect of species interactions can vary across space. Finally, *Z* is a normalising constant that cannot be computed for combinatorial reasons, although the statistical inference procedure takes care of it. Full details of the estimation procedure and parameter identifiability are available in Appendix S1: Section S.3 and Appendix S1: Section S.2.1, respectively.

Lastly, it is important to note two specificities of the metanetwork G^* in our modelling procedure: it cannot be directed nor contain selfloops. Indeed, Markov random fields specify conditional dependencies between random variables $\{X_i^l\}$ in an undirected way, and self-loops have no meaning in this framework. Our model assumes that these dependencies are given by the interaction network without considering the direction of edges. Consequently, this statistical model of interaction cannot be read in the light of causality. In case of trophic interactions, it consists in assuming that presence/absence of a predator and its prey are intertwined, without specifying top-down or bottomup control. Moreover, the absence of self-loops prevents from taking into account intraspecific effects. These effects are simply ignored by ELGRIN, as they are in any joint species distribution model or ordination technique (see Appendix S1: Section S.6).

ELGRIN is implemented in C++ for efficiency and is available in the function elgrin of the R package econetwork available on the code repository https://plmlab.math.cnrs.fr/econetproject/econetwork and at CRAN (https://cran.r-project.org/). We assessed the performance of the method in inferring parameters from data sampled and re-sampled under the model (see Appendix S1: Section S.4).

Table 1

Definition of variables and parameters of the Markov random field model ELGRIN.

Variables	Ecological interpretation
G^{\star}	Metanetwork of known interactions (undirected)
V^{\star}	Species (node set) of the metanetwork
E^{\star}	Interactions (edge set) of the metanetwork
X_i^l	Presence/absence of species i at location l
W	Environmental covariates at location l
Parameters	
a _i	Species <i>i</i> intercept
a_l	Location <i>l</i> intercept
b_i, c_i	Environmental (abiotic) parameters of species i
$\beta_{l,co-pres}$	Co-presence strength (or avoidance when < 0) at location l
$\beta_{l,co-abs}$	Co-absence strength (or avoidance when < 0) at location l

Model interpretation In the hypothetical example where G^* is an empty graph (no edges, none of the species interact), the random variables $\{X_i^l\}_{i \in V^*}$ are independent and each species is present with probability $e^{\alpha_{i,l}}/(1+e^{\alpha_{i,l}}) \in (0, 1)$, where $\alpha_{i,l} = a_l + a_i + W_l^{\mathsf{T}} b_i + (W_l^2)^{\mathsf{T}} c_i$. In other words, $\alpha_{i,l}$ is the logit of the probability of presence of species *i* at location *l* in the absence of interactions. Assuming that we have included all important environmental covariates, that there is no other ecological processes involved, and no model mis-specifications, $\alpha_{i,l}$ is analogous to the fundamental niche parameters of the species (sensu Hutchinson, 1959). It gives the probability of presence of species *i* at location *l* when only environmental filtering occurs.

In the case of species interactions, G^{\star} is a non empty graph and the presence/absence information is smoothed across neighbouring nodes in G^* . In Table 2, we detailed the ways both $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ parameters capture how the metanetwork influences species cooccurrences in a given location, notably the co-presence or co-absence of pairs of interacting species. This table describes expected patterns of species distribution according to the combination of positive, negative and zero values for the β parameters. More precisely, when species are known to interact positively (e.g. G^{\star} encodes mutualism) and that these interactions, averaged over all species with suitable environmental conditions at location *l*, influence their co-occurrences at that location, $\beta_{l,co-pres}$ and/or $\beta_{l,co-abs}$ will be estimated as positive. On the other hand, in case of negative interactions (e.g. G^{\star} encodes competition) that influence the co-occurrences at location l of species with favourable environmental conditions, the parameters $\beta_{l,co-pres}$ and/or $\beta_{l,co-abs}$ will be negative, co-presence configurations (or co-absence, respectively) tend to be avoided, meaning that only one of the two species tends to be present. Given a location with fixed total number of interacting co-present (resp. interacting co-absent) species, the larger the absolute value of $\beta_{l,co-pres}$ (resp. $\beta_{l,co-abs}$), the stronger the strength of the interactions.

2.3 Exploration on simulated data from complex dynamic processes

To test the ability of ELGRIN to infer the overall biotic and abiotic controls on species distributions, we used three theoretical models, different from the one underlying ELGRIN, to dynamically simulate spatial community data with 50 species and 400 sites along a single environmental gradient and combined them with multiple different interactions scenarios (competition, mutualism, and no interaction). To do that, we chose species niche optima evenly distributed along a single environmental gradient. The metanetworks were built so that interacting species have close niche optima (otherwise they would never co-occur). In the mutualistic scenario, we also considered a case where species that facilitate each other tend to have an abiotic niche that is also not too close (otherwise they would compete). Along this single environmental gradient, niche optima and associated metanetworks according the interaction scenarios, we used three theoretical dynamic models (Lotka-Volterra, colonisation-extinction, and co-existence model aka VirtualCom) to simulate the resulting species

distribution data. These models have different underlying assumptions and processes, which allowed testing ELGRIN under a total of 9 different configurations.

Lotka–Volterra model The Lotka–Volterra model is one of the foundational models in community ecology (Takeuchi, 1996). This model simulates communities under both intra- and interspecific interactions, while ELGRIN is not able to handle intraspecific interactions (its metanetwork does not allow for self-loops). Thus we parameterised the Lotka–Volterra simulation with intraspecific interactions being negligible in regards to interspecific interactions. That way we generated species community data that meets the type of data and ecological questions ELGRIN is designed to tackle (for details, see Appendix S1: Section S.5.1). Nonetheless, we also explored the converse case to fully understand the limits of ELGRIN (see Appendix S1: Section S.6).

Colonisation–extinction model We used an updated version of the stochastic colonisation–extinction model developed in Ohlmann et al. (2022) to simulate the species community dataset for the three interaction scenarios (for details see Appendix S1: Section S.5.2). The model consists in a multivariate Markov chain that converges towards a stationary distribution from which we sampled the species community dataset.

VirtualCom model We used an updated version of the model developed by Münkemüller and Gallien (2015) to simulate communities whose composition is driven simultaneously by biotic and abiotic environmental effects, for the three interaction scenarios (for details see Appendix S1: Section S.5.3). In this model, each community has the same carrying capacity (i.e. the exact number of individuals in each location).

2.4 Application: a case study

We analyse the newly available Tetra-EU 1.0 database, a specieslevel trophic network of European tetrapods (Maiorano et al., 2020) that combines all known potential interactions between terrestrial mammals, birds, reptiles and amphibians occurring in Europe. This metanetwork is based on data extracted from known interactions, scientific literature, including published articles, books, and grey literature (see Maiorano et al., 2020, for a complete description of the data and the reference list used to build the metanetwork). As usual with such data, this metanetwork does not provide information on interaction plasticity or intraspecific interactions. We restricted our analyses on the European Alps that show sharp environmental gradients and varying trophic web distributions (O'Connor et al., 2020). We extracted the species distribution data from Maiorano et al. (2013) at a 300 m resolution. We upscaled all species ranges maps to a 10×10 km equal-size area grid and cropped the distribution data to the European Alps. Species were considered present on a given 10×10 km cell if they were present in at least one of the 300 \times 300 m cells within it. This yielded species distributions maps for 257 breeding birds, 99 mammals, 36 reptiles, and 30 amphibians over 2138 locations. Environmental covariates were extracted at the same resolution and were selected following previous work on those data (Braga et al., 2019). For climate, we used mean annual temperature, temperature seasonality, temperature annual range, total annual precipitation and coefficient of variation of precipitation that were all extracted from the Worldclim v2 database (http://www.worldclim.org/ bioclim). Using GlobCover (GlobCover V2.2; http://due.esrin.esa.int/ page globcover.php), we extracted the number of habitats present in a given pixel, habitat diversity in a given pixel based on Simpson index and habitat evenness as a measure of habitat complexity. Finally, we added an index of annual net primary productivity (Global Patterns in Net Primary Productivity, v1 (1995), http://sedac.ciesin.columbia. edu/data/set/hanpp-net-primary-productivity) and the human footprint index (http://sedac.ciesin.columbia.edu/data/set/wildareas-v2human-footprint-geographic). Since these data were highly correlated, we used a PCA to retain the three leading vectors as environmental covariates (W_l) in ELGRIN.

Table 2

Simplified view of the different behaviours of the model in function of the parameters $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$. The graph represents the metanetwork containing all potential interactions where species can be either present (grey node) or absent (white node) in a given location *l* leading to different estimated $\beta_{l,co-abs}$ and $\beta_{l,co-abs}$. When $\beta_{l,co-abs} \ll 0$ or $\beta_{l,co-abs} \ll 0$, interacting species in the metanetwork tend to avoid each other: whenever one is absent, the other tend to be present and reversely. This situation favours a checkerboard pattern on the metanetwork. Reversely, whenever $\beta_{l,co-pres} \gg 0$ (resp. $\beta_{l,co-abs} \gg 0$), there are groups of interacting species that tend to be all present (resp. all absent), inducing sets of grey (resp. white) neighbour nodes in the metanetwork. Whenever $\beta_{l,co-pres} = 0$ or $\beta_{l,co-abs} = 0$, there are sets of interacting species whose states are independent from one another and thus purely random (the proportions of grey and white nodes are governed by the values of the parameters in the Grinnellian part of the model).



3 Results

3.1 Tests on simulated species community data

Let us first recall that we assessed the performance of the method in inferring parameters from data sampled and re-sampled under ELGRIN model (see Appendix S1: Section S.4). We now turn to dynamical theoretical ecological models.

For the three theoretical models (Lotka-Volterra, colonisationextinction and VirtualCom), ELGRIN was correct in identifying the no interaction scenario, with estimated interaction strengths close to 0 (Figs. 2-4). Similarly, ELGRIN was able to retrieve the negative effects of interactions in the case of competition as simulated by the three theoretical models. The $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ parameters were mostly negative (with much higher absolute values for $\beta_{l,co-pres}$), capturing the backbone of the competitive interactions. They indicated that co-presence and co-absence were avoided (as presented in Table 2 topleft), leading to some level of competitive exclusion. In the VirtualCom co-existence model, this phenomenon was clearly the by-product of the competitive interactions and the carrying capacity in terms of number of individuals (that explicitly induced exclusion). When positive interactions come into play (i.e. mutualism), the results should be contrasted between those obtained for the Lotka-Volterra model, where ELGRIN does not qualitatively identify the processes at stake and the two other models (colonisation-extinction and VirtualCom) where ELGRIN succeeds in identifying them. The Lotka-Volterra simulation with positive interactions scenario produced species that are essentially distributed along their respective niches (see Appendix S1: Figure S.4). As a consequence, this distribution can be simply fitted with the Grinellian part of the model and ELGRIN estimates the β s close to zero (Fig. 2). That means that the same dataset could have been produced by only abiotic environmental conditions and the actual species

distribution does not contain anymore a pattern that ELGRIN would identify as the trace of the positive interspecific interactions. On the contrary, in the positive interactions scenario, with both competitioncolonisation and VirtualCom co-existence models, ELGRIN correctly identified the process at play. The parameters $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ were mostly positive. During the simulation steps, the presence of one species was then favoured by the presence of another species it interacted with, leading to a co-presence phenomena captured by the positive $\beta_{l,co-pres}$. Conversely, the inverse mechanism emerged for coabsence, implying that the $\beta_{l,co-abs}$ tended to be positive as revealed by ELGRIN (Figs. 3, 4). To quantitatively investigate the difference between $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ distributions in the three simulations, we performed Kolmogorov-Smirnov (KS) tests. For each simulation, we tested whether $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ distributions were significantly different in the scenarios with interactions (either positive or negative) from the scenario without interaction. In the three simulations, the tests correctly identify significant differences between interactions and no interaction scenarios (see Appendix S1: Table S.1).

3.2 Empirical case study

When fitted to the European vertebrate dataset, ELGRIN's parameters $\beta_{l,co-pres}$ and $\beta_{l,co-abs}$ were highly correlated (Pearson correlation of 0.84, see Appendix S1: Section S.7.1) suggesting that trophic interactions impact both predator/prey co-presence and co-absence. In what follows, we therefore mainly dealt with $\beta_{l,co-pres}$.

We first observed a structured spatial pattern of the effects of interactions, with regions of negative or positive $\beta_{l,co-pres}$ (bluish or reddish colours respectively in Fig. 5). The largest $\beta_{l,co-pres}$ values were found mainly in the french Alps and in the Eastern zone.

In Fig. 6, we present the values of different variables at each location, according to groups of estimated $\beta_{l,co-pres}$ parameters, where



Fig. 2. Distribution of co-presence ($\beta_{l,co-pres}$) and co-absence ($\beta_{l,co-abs}$) strengths inferred using ELGRIN on simulated ecological communities using a Lotka–Volterra model with competition (negative interactions), mutualism (positive interactions) or no interactions.



Fig. 3. Distribution of co-presence ($\beta_{l,co-pres}$) and co-absence ($\beta_{l,co-abs}$) strengths inferred using ELGRIN on simulated ecological communities using a colonisation-extinction model with competition (negative interactions), mutualism (positive interactions) or no interactions.

the width of each boxplot is proportional to the number of points in each class. Almost all the highest $\beta_{l,co-pres}$ (> 0.05) were revealed in locations below 1600 m of altitude (Fig. 6a, p-value of the KS test inferior to 2.2e-16, details given in Appendix S1: Section S.7.2). In these regions, species richness was generally high (Fig. 6b, p-value inferior to 2.2e–16). In the opposite, the higher up, the more likely $\beta_{l,co-pres}$ was negative (Fig. 6a). This was particularly true above 1600 m in the central Alps, where almost all the negative $\beta_{l,co-pres}$ were estimated (bluish colours in Fig. 5). Locations with negative $\beta_{l,co-pres}$ have a lower species richness (Fig. 6b). Interestingly, locations with low connectance have lower absolute $\beta_{l,co-pres}$ values (Fig. 6c, *p*-value inferior to 2.2e–16) indicating a lower effect of biotic interactions compared to abiotic effects in these locations. Here, connectance is the density of the graph induced by the metanetwork at location *l*, namely its nodes are species occurring at location l and edges are those from the metanetwork between those present species.

4 Discussion

Deciphering the mechanisms driving spatial patterns of species distributions and communities is likely one of the most active fields of ecological research since the early days of biogeography and community ecology. Still, there was so far no comprehensive statistical approach able to make the best of existing knowledge on interspecific interactions, species occurrence and environmental data to measure and quantify the dual effects of environment and biotic interactions on species distributions. Our proposed model that relies on Markov random fields builds on the ability of graphical models to encode and analyse species distribution dependencies using the known species interactions. This formalism allows, within the same model, to account for both the effects of the environment and the interspecific interactions, which reconciles the Grinnellian vision of species niches (i.e. how species respond to the abiotic environment) with its Eltonian



Fig. 4. Distribution of co-presence ($\rho_{l,co-pres}$) and co-absence ($\rho_{l,co-abs}$) strengths inferred using ELGRIN on simulated ecological communities with VirtualCom model, with competition (negative interactions), mutualism (positive interactions) or no interactions.



Fig. 5. Results of ELGRIN on the European tetrapods case study. Map of estimated $\beta_{l,co-pres}$ (one dot per location). The colour scale indicates the $\beta_{l,co-pres}$ values. For the sake of representation, $\beta_{l,co-pres}$ values above 0.15 in absolute value were set to 0.15.

counterpart (i.e. how species respond to the biotic environment). The mathematical foundations of ELGRIN are strong and its framework is flexible allowing for useful extensions to handle interaction strength, sampling effects and plasticity of interactions (see Appendix S1: Section S.1).

A key element of ELGRIN is its ability to measure the overall relative effects of interspecific interactions on species distributions with respect to abiotic environmental conditions, which allows to summarise all local pairwise interactions in a single measure (i.e. $\beta_{l,co-abs}$ or $\beta_{l,co-pres}$). This measure can then be mapped, related to spatial layers to understand how the overall relative effect of interspecific interactions vary in space and in function of the environment or the ecosystem types. Importantly, this measure can also be carefully investigated at a given location in function of the constituent species, trophic groups, specialists vs generalists, connectance and so on. Interestingly, we can thus see our β_l estimates as an extended and more meaningful version

of the famous checkerboard score or C-score (Stone and Roberts, 1990), which has been used to quantify local interspecific interactions from cooccurrence pattern (e.g., Boulangeat et al., 2012). The main advantage of ELGRIN over the C-score is that instead of trying to infer biotic interactions only from co-occurrences (which we know to be notoriously difficult, nearly impossible), it quantifies, in a conditional way, the effects of the known interspecific interactions on species communities, while accounting for the environmental responses of the species. Our approach is thus not comparable with recent developments on joint species distribution models (JSDMs) that relate species occurrences to environmental conditions, and provides a residual covariance matrix that could be interpreted on the light of missing predictors, misspecifications and biotic interactions (Ovaskainen et al., 2017; Zurell et al., 2018). This matrix represents covariances between model residuals (the left-over from the environmental effects) and actually provides little information about biotic interactions (Zurell et al., 2018; Poggiato



Fig. 6. Results of ELGRIN on the European tetrapods case study. Boxplots representing the values of different variables at each location, according to the estimated $\beta_{l,co-pres}$ values (x axis). (a) altitude, (b) species richness, and (c) connectance (density of the graph induced by the metanetwork at location *l*) For the sake of representation, $\beta_{l,co-pres}$ values above 0.15 in absolute value were set to 0.15. Width of the boxplots is proportional to the number of points in each class.

et al., 2021). On the contrary, ELGRIN does not infer any residual covariance and directly accounts for the known interactions through the metanetwork. In JSDMs, missing covariates will inevitably lead to spurious estimates of biotic interactions. In ELGRIN, the parameter a_l is supposed to capture most of the unexplained information that is independent of the interspecific interactions. This parameter acts as a site random effect in mixed models and is expected to filter out the effects of missing covariates, although some remaining species-specific effects might still percolate into the β_l estimates.

In the presentation of ELGRIN and in our case studies, we focused on a single interaction type at a time (e.g. competition, mutualism or trophic interaction). When dealing with a single type of interaction, competition for instance, the modelling is explicit since we clearly understand the effect that one species can have on another species. Although it is technically possible to manage a metanetwork composed of different types of interactions, the interpretation would become problematic. Different interaction types can have opposite effects, such as competition (a species excludes other species) and mutualism (a species facilitates other species) and, since ELGRIN captures an overall impact of these interactions on the distributions at each location, interpreting ELGRIN's results can be misleading in that case. Additionally, it is worth noting that since ELGRIN relies on a Markov random field, G^{\star} is undirected. In other words, when the original metanetwork encodes asymmetric interactions (e.g. predator-prey), they are then converted in undirected edges that only represent the presence of interactions (whatever their direction). It is thus critical to keep that in mind when interpreting the results of ELGRIN, and when merging different types of interactions together. The same issue happens when hoping to interpret the residual covariance matrix of JSDM through the lens of biotic interactions, since the values of the covariance matrix could reflect any type of interactions between species, that could be asymmetric or symmetric, or both. Note that we explicitly used a bell-shaped relationship for modelling species response to environmental gradients. While it would be possible to modify ELGRIN to incorporate any other parametric relationship, the actual version of ELGRIN would lead to

erroneous conclusions whenever used on data where this assumption is not satisfied.

More generally, it is important to underline that ELGRIN finds the most likely scenario under a model associated to underlying assumptions. This model represents up to date the most reasonable and simple model that integrates both interspecific interactions and abiotic factors in modelling the species distribution. In that sense, it goes beyond (joint) species distribution models or ordination models by including explicitly the effect of interspecific interactions. However, the most likely scenario under this model is not necessarily the real one that lead to observed data. For instance, ELGRIN was not able to identify the positive interspecific interactions present in the dynamics of a Lotka-Volterra model (even when restricting to negligible intraspecific interactions). Despite being a most widely studied model, the Lotka-Volterra model still raises important challenges. Indeed, whether the system reaches a single globally stable equilibrium point is known only in specific cases (Takeuchi, 1996). Since ELGRIN infers model interspecific interactions relative effects from the species distributions, existence of multiple equilibria in the Lotka-Volterra dynamics (depending on the initial conditions that are unknown) could pose serious identifiability problems. Even in presence of a unique and globally stable equilibrium point, several parameters or different interaction types could lead to the same equilibrium and thus same observed species distributions. This also raises tough identifiability issues. We hope that the recent developments around Lotka-Volterra model will help to circumvent those issues (Biroli et al., 2018; Remien et al., 2021). We could easily simulate species distributions, using models that include other ecological processes, on which ELGRIN would fail in recovering the true underlying generation processes. Indeed we present simulations scenarios beyond the assumptions of the model (i.e., a Lotka-Volterra model with intraspecific interactions stronger than interspecific ones, see Appendix S1: Section S.6), where ELGRIN again uncovered a completely different explanation of the data at hand. If the data contain the signature of different ecological processes (including ones not considered by ELGRIN), ELGRIN will not be able to infer properly the relative effects of interspecific interactions and abiotic factors. The question of knowing which ecological processes could indeed be recovered from species distribution patterns remains thus debated (e.g. Blanchet et al., 2020). A last note is that ELGRIN only deals with binary occurrence data rather than abundance or frequency data. In our simulation design, both the Lotka–Volterra and the Virtual-Com models produced abundance data that we had to sample to obtain binary signals, losing information during the process. On the contrary, ELGRIN performs better on colonisation–extinction simulations, where the dynamics directly generates binary data. Extending ELGRIN from the binary setup to the continuous one could improve the inference by considering more information in the species distribution data but it remains an important methodological challenge.

In terms of further perspectives, we might wonder whether this model could be extended for prediction purposes. In principle, it is possible to draw presence/absence data from the model for different values of the environment variables. These different values could allow for predictions in space but also in time. However, something to keep in mind is that the metanetwork will not vary in the model and will thus be considered as static and thus representative in space (or in time). If the metanetwork has not been built with that prediction perspective in mind, this might be an issue as we will miss interaction rewiring effects on species distributions. Instead, if the metanetwork is truly a potential metanetwork that tries to incorporate these potential interactions that have been observed yet (i.e. Maiorano et al., 2020), it might be interesting to investigate how biotic interactions might further influence future species distributions in response to environmental changes.

CRediT authorship contribution statement

Marc Ohlmann: Performed the simulations and focus on their interpretation, Writing of the manuscript, Revisions, Designed the applications. **Catherine Matias:** Developed a first version of the statistical framework, Writing of the manuscript, Revisions, Designed the applications. **Giovanni Poggiato:** Performed the simulations and focus on their interpretation, Writing of the manuscript, Revisions, Designed the applications. **Stéphane Dray:** Provided critical perspectives on the model and the applications, Writing of the manuscript, Revisions, Designed the applications. **Wilfried Thuiller:** Provided critical perspectives on the model and the applications, Writing of the manuscript, Revisions, Designed the applications, Writing of the manuscript, Revisions, Designed the applications. **Vincent Miele:** Implemented the applications, Writing of the manuscript, Revisions, Writing of the manuscript, Revisions, Writing of the manuscript, Revisions, Designed the applications. **Vincent Miele:** Implemented the applications, Writing of the manuscript, Revisions, Designed the applications. **Vincent Miele:** Implemented the applications, Writing of the manuscript, Revisions.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Catherine Matias reports financial support was provided by French National Research Agency.

Data availability

We used a open dataset already available online and shared the access link in the manuscript

Acknowledgements

The authors would like to thanks the anonymous reviewers that carefully reviewed a previous version of our model and manuscript, and notably proposed the Lotka–Volterra simulation scheme. Funding was provided by the French National Center for Scientific Research (CNRS) and the French National Research Agency (ANR) grant ANR-18-CE02-0010-01 EcoNet. VM would like to thank the LECA laboratory for hosting him in Chambéry.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ecolmodel.2023.110424.

References

- Austin, M., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. 157 (2), 101–118.
- Biroli, G., Bunin, G., Cammarota, C., 2018. Marginally stable equilibria in critical ecosystems. New J. Phys. 20 (8), 083051.
- Blanchet, F.G., Cazelles, K., Gravel, D., 2020. Co-occurrence is not evidence of ecological interactions. Ecol. Lett. 23 (7), 1050–1063.
- Botella, C., Dray, S., Matias, C., Miele, V., Thuiller, W., 2022. An appraisal of graph embeddings for comparing trophic network architectures. Methods Ecol. Evol. 13 (1), 203–216.
- Boulangeat, I., Gravel, D., Thuiller, W., 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecol. Lett. 15 (6), 584–593.
- Braga, J., Pollock, L.J., Barros, C., Galiana, N., Montoya, J.M., Gravel, D., Maiorano, L., Montemaggiori, A., Ficetola, G.F., Dray, S., et al., 2019. Spatial analyses of multitrophic terrestrial vertebrate assemblages in Europe. Global Ecol. Biogeogr. 28 (11), 1636–1648.
- Brémaud, P., 1999. Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues, Vol. 31. Springer.
- de Candolle, A., 1855. Géoographie Botanique Raisonnée ou, Exposition des Faits Principaux Et Des Lois Concernant la Distribution Géographique Des Plantes de L'époque Actuelle. Masson.
- Chalmandrier, L., Münkemüller, T., Gallien, L., De Bello, F., Mazel, F., Lavergne, S., Thuiller, W., 2013. A family of null models to distinguish between environmental filtering and biotic interactions in functional diversity patterns. J. Veg. Sci. 24 (5), 853–864.
- Chase, J.M., Leibold, M.A., 2003. Ecological Niches: Linking Classical and Contemporary Approaches. University of Chicago Press.
- Cirtwill, A.R., Eklöf, A., Roslin, T., Wootton, K., Gravel, D., 2019. A quantitative framework for investigating the reliability of empirical network construction. Methods Ecol. Evol. 10 (6), 902–911.
- Connor, E.F., Simberloff, D., 1979. The assembly of species communities: chance or competition? Ecology 60 (6), 1132–1140.
- Diamond, J.M., 1975. Assembly of species communities. In: Diamond, J., Cody, M. (Eds.), Ecology and Evolution of Communities. Harvard University Press, pp. 342–444.
- Gravel, D., Baiser, B., Dunne, J.A., Kopelke, J.-P., Martinez, N.D., Nyman, T., Poisot, T., Stouffer, D.B., Tylianakis, J.M., Wood, S.A., Roslin, T., 2019. Bringing elton and grinnell together: a quantitative framework to represent the biogeography of ecological interaction networks. Ecography 42 (3), 401–415.
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models: With Applications in R. In: Ecology, Biodiversity and Conservation, Cambridge University Press.
- Holt, R.D., 2020. Some thoughts about the challenge of inferring ecological interactions from spatial data. Biodivers. Inform. 15 (1), 61–66.
- Hutchinson, G.E., 1959. Homage to Santa Rosalia or why are there so many kinds of animals? Amer. Nat. 93 (870), 145–159.
- Lortie, C.J., Brooker, R.W., Choler, P., Kikvidze, Z., Michalet, R., Pugnaire, F.I., Callaway, R.M., 2004. Rethinking plant community theory. Oikos 107 (2), 433–438.
- Maiorano, L., Amori, G., Capula, M., Falcucci, A., Masi, M., Montemaggiori, A., Pottier, J., Psomas, A., Rondinini, C., Russo, D., et al., 2013. Threats from climate change to terrestrial vertebrate hotspots in europe. PLoS One 8 (9).
- Maiorano, L., Montemaggiori, A., O'Connor, L., Ficetola, G., W., T., 2020. TETRA-EU 1.0: A species-level trophic meta-web of European tetrapods. Glob. Ecol. Biogeogr. 29 (9), 1452–1457.
- Morales-Castilla, I., Matias, M.G., Gravel, D., Araújo, M.B., 2015. Inferring biotic interactions from proxies. Trends Ecol. Evol. 30 (6), 347–356.
- Münkemüller, T., Gallien, L., 2015. VirtualCom: a simulation model for eco-evolutionary community assembly and invasion. Methods Ecol. Evol. 6 (6), 735–743.
- O'Connor, L.M.J., Pollock, L.J., Braga, J., Ficetola, G.F., Maiorano, L., Martinez-Almoyna, C., Montemaggiori, A., Ohlmann, M., Thuiller, W., 2020. Unveiling the food webs of tetrapods across Europe through the prism of the Eltonian niche. J. Biogeogr. 47 (1), 181–192.
- Ohlmann, M., Miele, V., Dray, S., Chalmandrier, L., O'Connor, L., Thuiller, W., 2019. Diversity indices for ecological networks: a unifying framework using hill numbers. Ecol. Lett. 22 (4), 737–747.
- Ohlmann, M., Munoz, F., Massol, F., Thuiller, W., 2022. Assessing Mutualistic Metacommunity Capacity by Integrating Spatial and Interaction Networks. Technical Report, arXiv:2206.11029.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, F.G., Duan, L., Dunson, D., Roslin, T., Abrego, N., 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol. Lett. 20 (5), 561–576.

- Pellissier, L., Albouy, C., Bascompte, J., Farwig, N., Graham, C., Loreau, M., Maglianesi, M.A., Melián, C.J., Pitteloud, C., Roslin, T., et al., 2018. Comparing species interaction networks along environmental gradients. Biol. Rev. 93 (2), 785–800.
- Peres-Neto, P.R., Olden, J.D., Jackson, D.A., 2001. Environmentally constrained null models: site suitability as occupancy criterion. Oikos 93 (1), 110–120.
- Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J.S., Thuiller, W., 2021. On the interpretations of joint modeling in community ecology. Trends Ecol. Evol. 36 (5), 391–401.
- Pulliam, H., 2000. On the relationship between niche and distribution. Ecol. Lett. 3, 349–361.
- Remien, C.H., Eckwright, M.J., Ridenhour, B.J., 2021. Structural identifiability of the generalized Lotka–Volterra model for microbiome studies. R. Soc. Open Sci. 8 (7), 201378.
- Ricklefs, R.E., 2008. Disintegration of the ecological community: American Society of Naturalists Sewall Wright Award Winner Address. Amer. Nat. 172 (6), 741–750.
- Soberón, J., Nakamura, M., 2009. Niches and distributional areas: Concepts, methods, and assumptions. Proc. Natl. Acad. Sci. 106 (supplement_2), 19644–19650.

- Staniczenko, P.P., Sivasubramaniam, P., Suttle, K.B., Pearson, R.G., 2017. Linking macroecology and community ecology: refining predictions of species distributions using biotic interaction networks. Ecol. Lett. 20 (6), 693–707.
- Stone, L., Roberts, A., 1990. The checkerboard score and species distributions. Oecologia 85 (1), 74–79.
- Takeuchi, Y., 1996. Global Dynamical Properties of Lotka-Volterra Systems. World Scientific, Singapore.
- Thuiller, W., Pollock, L.J., Gueguen, M., Münkemüller, T., 2015. From species distributions to meta-communities. Ecol. Lett. 18 (12), 1321–1328. http://dx.doi.org/10. 1111/ele.12526.
- Tikhonov, G., Abrego, N., Dunson, D., Ovaskainen, O., 2017. Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. Methods Ecol. Evol. 8 (4), 443–452.
- Tylianakis, J.M., Morris, R.J., 2017. Ecological networks across environmental gradients. Annu. Rev. Ecol. Evol. Syst. 48, 25–48.
- Zurell, D., Pollock, L.J., Thuiller, W., 2018. Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? Ecography 41 (11), 1812–1819.